INFORMATION AND CONTROL 6, 126–130 (1963)

A Note on "Very Noisy" Channels

BARNEY REIFFEN

Lincoln Laboratory, Massachusetts Institute of Technology*
Lexington, Massachusetts

A "very noisy" channel is defined. This definition corresponds to many physical channels operating at low signal-to-noise ratio. For "very noisy" discrete input memoryless channels, the computation cutoff rate for sequential decoding, R_{comp}, is shown to be one-half the capacity, C. Furthermore, that choice of input probabilities which achieves C also maximizes R_{comp}, and vice versa.

I. INTRODUCTION

A memoryless channel is defined by:

(a) an input alphabet X,
(b) an output alphabet Y,
(c) a transition probability density $p(y/x)$.

For any probability density $p(x)$ defined on the input space, the density on the product space XY is well defined, and the random variable mutual information $I(x; y)$ may be defined:

$$I(x; y) = \ln \frac{p(y \mid x)}{p(y)} \tag{1}$$

where

$$p(y) = \int p(x)p(y \mid x)\, dx \tag{2}$$

The average of the mutual information, denoted by $I(X; Y)$, can be written

$$I(X; Y) = H(Y) - H(Y|X) \tag{3}$$

* Operated with support from the United States Army, Navy, and Air Force.

where

$$H(Y) = -\int p(y) \ln p(y)\, dy \tag{4}$$

$$H(Y \mid X) = -\iint p(x)p(y \mid x) \ln p(y \mid x)\, dx\, dy \tag{5}$$

In this formulation, the Y space is assumed to be continuous. The X space may be continuous or discrete. In the latter case $p(x)$ contains delta functions. The results of this note are trivially extended to discrete input-discrete output channels.

In order to obtain the desired results, we define a channel as "very noisy" if for all $x \in X$

$$\frac{p(y) - p(y \mid x)}{p(y)} = \epsilon_x(y) \ll 1. \tag{6}$$

This description corresponds to many physical channels operating at low signal-to-noise ratio.

Observe that

$$\int p(x)\epsilon_x(y)\, dx = 0 \tag{7}$$

and

$$\int p(y)\epsilon_x(y)\, dy = 0 \tag{8}$$

II. EVALUATION OF C FOR "VERY NOISY" CHANNELS

We first evaluate $H(Y|X)$.

$$\begin{aligned} H(Y \mid X) &= -\int p(x) \left[\int p(y \mid x) \ln p(y \mid x)\, dy\right] dx \\ &= -\int p(x) \left[\int dy p(y)[1 - \epsilon_x(y)] \right. \\ &\qquad \left. \times \ln \{p(y)[1 - \epsilon_x(y)]\}\right] dx \end{aligned} \tag{9}$$

Using the inequality

$$\ln(1 + \alpha) \geqq \alpha - \tfrac{1}{2}\alpha^2, \qquad \alpha^2 < 1 \tag{10}$$

the bracketed expression of Eq. (9) reduces to

$$
\begin{aligned}
[\quad] \geqq \int p(y) \ln p(y)\, dy - \int p(y)\epsilon_x(y) \ln p(y)\, dy \\
- \int p(y)\epsilon_x(y)\, dy + \frac{1}{2} \int p(y)\epsilon_x^2(y)\, dy + \frac{1}{2} \int p(y)\epsilon_x^3(y)\, dy
\end{aligned}
\tag{11}
$$

In Eq. (11), the first integral is $-H(Y)$. The third integral is zero (Eq (8)). The fifth integral will be neglected, since it is a third order term in $\epsilon_x(y)$ which is assumed very much less than unity. Combining these observations, we have, to within terms of second order in $\epsilon_x(y)$,

$$
\begin{aligned}
H(Y \mid X) = H(Y) + \iint p(x)p(y)\epsilon_x(y) \ln p(y)\, dx\, dy \\
- \frac{1}{2} \iint p(x)p(y)\epsilon_x^2(y)\, dx\, dy
\end{aligned}
\tag{12}
$$

The first integral in Eq. (12) is zero (Eq. (7)). Thus, we have from Eqs. (3) and (12)

$$
I(X; Y) = H(Y) - H(Y \mid X) = \frac{1}{2} \iint p(x)p(y)\epsilon_x^2(y)\, dx\, dy \tag{13}
$$

Equation (13) is correct to within terms of second order of $\epsilon_x|(y)$.

The channel capacity C is defined as

$$
C = \max_{p(x)} I(X; Y) \tag{14}
$$

For that $p(x)$ which yields capacity,

$$
C = \frac{1}{2} \iint p(x)p(y)\epsilon_x^2(y)\, dx\, dx \tag{15}
$$

where Eq. (15) is correct to within terms of second order of $\epsilon_x(y)$.

III. EVALUATION OF R_{comp} FOR "VERY NOISY" CHANNELS

The quantity

$$
E_0 = -\ln \int \left[\int p(x)[p(y \mid x)]^{1/2}\, dx \right]^2 dy \tag{16}
$$

has considerable physical significance. E_0 is the exponent in the upper bound to error probability corresponding to zero rate. This has been shown by Fano (1961) for the case of discrete channels, but the generalization to discrete input-continuous output channels is straightforward.

For the sequential decoding procedure (Wozencraft and Reiffen, 1961; Reiffen, 1962), a computation cutoff rate, R_{comp}, is defined. For rates less than R_{comp}, the average number of decoding computations does not grow exponentially with constraint length, but algebraically. Reiffen (1962) has shown that for discrete input channels,

$$R_{\text{comp}} \leqq E_0 \tag{17}$$

It is conjectured in the reference that Eq. (17) is an equality rather than an inequality. This conjecture has been verified by Fano in unpublished work. Thus, restricting our attention to discrete input channels, we may write

$$R_{\text{comp}} = -\ln \int \left[\int p(x)[p(y \mid x)]^{1/2}\, dx \right]^2 dy \tag{18}$$

where it is understood that the density $p(x)$ contains delta functions.

Starting with Eq. (18) we obtain

$$\begin{aligned} \int \left[\int p(x)[p(y \mid x)]^{1/2}\, dx \right]^2 dy & \\ = \int \left[\int p(x)\{p(y)[1 - \epsilon_x(y)]\}^{1/2}\, dx \right]^2 dy & \qquad (19) \\ \geqq \int p(y) \left[\int p(x)[1 - \tfrac{1}{2}\epsilon_x(y) - \tfrac{1}{8}\epsilon_x^2(y)]\, dx \right]^2 dy & \end{aligned}$$

where we have used the inequality

$$(1 - \alpha)^{1/2} \geqq 1 - \tfrac{1}{2}\,\alpha - \tfrac{1}{8}\,\alpha^2, \qquad \alpha^2 < 1 \tag{20}$$

The bracketed expression on the right-hand side of Eq. (19) then becomes

$$[\quad] = \int p(x)\, dx - \frac{1}{2} \int p(x)\epsilon_x(y)\, dx - \frac{1}{8} \int p(x)\epsilon_x^2(y)\, dx \tag{21}$$

The second integral in Eq. (21) is zero (Eq. (7)). Thus,

$$[\quad]^2 = \left[1 - \frac{1}{8} \int p(x)\epsilon_x^2(y)\, dx \right]^2 \geqq 1 - \frac{1}{4} \int p(x)\epsilon_x^2(y)\, dx \tag{22}$$

Combining Eqs. (18), (19), and (22), we obtain

$$e^{-R_{\text{comp}}} \geqq 1 - \frac{1}{4} \iint p(x)p(y)\epsilon_x^2(y)\, dx\, dy \tag{23}$$

From Eq. (23) we conclude that, within second order terms of $\epsilon_x(y)$,

$$R_{\text{comp}} = \frac{1}{4} \iint p(x)p(y)\epsilon_x^2(y)\, dx\, dy \tag{24}$$

Comparing Eqs. (13) and (24), we see that for any $p(x)$,

$$R_{\text{comp}} = \tfrac{1}{2}\, I(X;Y) \tag{25}$$

Thus, that choice of $p(x)$ which maximizes $I(X;Y)$ and thereby achieves capacity also maximizes R_{comp}, and vice versa.

IV. CONCLUSIONS

For "very noisy" discrete input memoryless channels

(a) $R_{\text{comp}} = \frac{1}{2}$ C.

(b) That choice of input probabilities which achieves C also maximizes R_{comp}, and vice versa.

(c) The capacity may be evaluated indirectly by calculating R_{comp} and doubling it. This might be easier than a direct evaluation of C since $\exp(-R_{\text{comp}})$ is an algebraic function of $p(x)$, while $I(X; Y)$ is a transcendental function of $p(x)$.

Received: December 26, 1962

References

Fano, R. M. (1961), "Transmission of Information," p. 333. M.I.T. Press and Wiley, New York.

Wozencraft, J. M., and Reiffen, B. (1961), "Sequential Decoding." M.I.T. Press and Wiley, New York.

Reiffen, B. (1962), Sequential decoding for discrete input memoryless channels. *Trans. IRE, PGIT* **IT-8**, 208–220.